

Maximum entropy niche-based modeling (Maxent) of potential geographical distributions of fruit flies *Dacus bivittatus*, *D. ciliatus* and *D. vertebrates* (Diptera: Tephritidae)

LI Bai-Ni¹, WEI Wu¹, MA Jun², ZHANG Run-Jie^{1,*}

(1. State Key Laboratory for Biocontrol & Institute of Entomology, Sun Yat-Sen University, Guangzhou 510275, China;

2. Plant Quarantine Laboratory, Guangdong Inspection and Quarantine Technology Center, Guangzhou 510623, China)

Abstract: In order to predict and analyse the potential geographical distributions of three important quarantine invasive pests, *Dacus bivittatus*, *D. ciliatus* and *D. vertebrates*, three ecological niche modeling techniques, BIOCLIM, DOMAIN and Maximum entropy niche-based modeling (Maxent) were implemented by using distribution records of the three fruit fly species and a set of environmental predictor variables. Differences in prediction performance of the three models with thresholds were observed. An evaluation using independent records of *D. bivittatus* showed that Maxent offers the most accurate predictions than two other models based on three values of ROC/AUC, Kappa, and TSS. Prediction outcomes made by Maxent revealed that the three fruit fly species have broadly similar potential ranges in Central American, South American, Southeast Asia, and Coastal areas of Australia in general. *D. ciliatus* has the comparatively widest potential range among the three species, including Coastal areas of Mediterranean Sea, Saudi Arabia, Yemen, Oman and South Iran, suggesting that it may be tolerant of the widest range of climatic conditions among the three species. In China, while large areas of Yunnan and Hainan are very habitable for all the three fruit fly species, southern part of Guangdong and Taiwan are also their habitable areas. *D. ciliatus* has the widest potential distribution area, with southern part of Sichuan, Guizhou and Tibet plus the coastal areas of southern China all being its suitable areas. The risk of the three fruit fly species permanently establishing in Guangdong if introduced exist but low. Jackknife analysis revealed that temperature and its variation have comparative significant influence on the distribution patterns of three fruit fly species both in global and restricted regions.

Key words: *Dacus bivittatus*; *D. ciliatus*; *D. vertebrates*; niche modeling; maximum entropy niche; Maxent; BIOCLIM; DOMAIN

1 INTRODUCTION

Guangdong province is situated at southern China where the climate is suitable for fruit fly species, and is adjacent to Hong Kong. It is a major gateway for international commercial trade and traveler transportation in China mainland. About one third of total imported fruit consignments of the country enter China mainland through Guangdong ports each year. From 2003 to 2007, *Dacus bivittatus* (Bigot) and *D. ciliatus* Loew had been intercepted four and three times respectively in traveler inspection at Guangdong ports, and there are other five times of *Dacus* spp. interception records in addition (unpublished data) (Liang *et al.*, 2006, 2007). The Chinese Government classified *Dacus* spp. as the important quarantine invasive pests because of the economic and international business significance (AQSIQ, 2005; CABI, 2005). *D. bivittatus*, *D. ciliatus* and *D. vertebrates* Bezzi are three important and

serious invasive pest species of *Dacus* spp., whose native distribution areas are Africa (Liang *et al.*, 1993; CABI, 2005; Liang *et al.*, 2006; EPPO, 2007), and there are no distribution reports of them in China yet. Nowadays, with more and more frequent business trade between China and African countries, the interception frequency and potential risk of introduction of *Dacus* spp. into China are growing higher. Given the tropical and subtropical climatic features of southern China, and the wide areas of fruit, vegetable and forest there, it is very likely that the *Dacus* spp. pests cause serious damage to the crops and environmental security if they introduced (Liang *et al.*, 2006, 2007). Researches about predictions of potential distributions of *D. bivittatus*, *D. ciliatus* and *D. vertebrates* are few but necessary.

Therefore, objectives of this study were two: (1) Predicting and comparing potential geographical distributions of *D. bivittatus*, *D. ciliatus* and *D.*

基金项目: 国家“973”计划项目(2002CB111405); 国家自然科学基金项目(30471162, 30671394); 国家科技支撑计划项目(2006BAD08A13)

作者简介: 李白尼, 男, 1984年生, 硕士, 研究方向为生物入侵与控制, E-mail: bainilee@gmail.com

* 通讯作者 Author for correspondence, E-mail: lsszrj@mail.sysu.edu.cn

收稿日期 Received: 2009-02-12; 接受日期 Accepted: 2009-09-01

vertebratus, both in China and global area. The process was launched based on ecological niche requirements (largely in climatic dimensions) and their ecological distributions in their origin regions. Models of potential distribution are developed and tested quantitatively for *D. bivittatus* as a measure of the predictive power of the methodology, then applied to two others; (2) Finding the critical environmental variables determining the distribution patterns of the three fruit fly species.

2 MATERIALS AND METHODS

2.1 Occurrence data

Native-range distributional data for *D. bivittatus*, *D. ciliatus* and *D. vertebratus* were obtained from Royal Museum for Central Africa (<http://projects.bebif.be/enbi/fruitfly>). The database contains Afrotropical specimen data for the ceratitidine (Tephritidae, Dacinae, tribe Ceratitidini) fruit fly genera *Ceratitis*, *Trirhithrum* Bezzi, and *Capparimyia* Bezzi, and summarizes the holdings of all major European, African and North American natural history collections, supplemented with a few records from specimens housed in the National Insect Collection at Pretoria, South Africa (De Meyer *et al.*, 2008).

Non-native distributional data was gathered from literature, research papers, reports, and electronic resources. The data reflect both contemporary as well as historical records of the three species outside Africa, and permanent as well as temporary establishments.

Only records with authoritative confirmation of identification were included; those with doubtful identification were excluded from model building, besides we used only records of permanent establishment for distribution predicting. In total, 345, 343 and 243 records were adopted for *D. bivittatus*, *D. ciliatus* and *D. vertebratus*, respectively.

Localities were georeferenced based either on coordinates provided directly on specimen labels for individual specimens, or on coordinates extracted from official gazetteers (*e. g.*, GeoNet, <http://gnswww.nga.mil/geonames/GNS/index.jsp>). Records with unspecified or unknown localities were removed from consideration. In cases of multiple records for a particular locality, only one was used in analysis.

2.2 Environmental data sets and relevant geographic layers

The environmental layers ‘bioclimatic’ were used in predicting procedure, which consist of important ecological factors and climatic features worldwide. It contains 19 layers with 2.5° spatial resolution, and was derived by the WorldClim project, <http://biogeo.berkeley.edu>. These were Annual Mean Temperature, Annual Precipitation, Isothermality, Max Temperature of Warmest Month, Mean Diurnal Range, Mean Temperature of Coldest Quarter, Mean Temperature of

Driest Quarter, Mean Temperature of Warmest Quarter, Mean Temperature of Wettest Quarter, Min Temperature of Coldest Month, Precipitation of Coldest Quarter, Precipitation of Driest Month, Precipitation of Driest Quarter, Precipitation of Warmest Quarter, Precipitation of Wettest Month, Precipitation of Wettest Quarter, Precipitation Seasonality, Temperature Annual Range and Temperature Seasonality. Procedures producing these interpolated climate surfaces are described by Hijmans *et al.* (2005).

Layers of the 1:4 million-scale about administrative boundary of China were downloaded from National Fundamental Geographic Information System (NFGIS: <http://nfgis.nsdi.gov.cn/>). The layers characterized the geographic location of provinces boundary, main rivers, important cities, large habitat areas, main road lines and main railway lines of China. We covered them on the final prediction outputs to analyse the distribution patterns of three fruit fly species.

No vegetation-cover layers and architectural composition layers were used, because they change markedly year by year; and besides given the heterogeneous nature of habitats (including man-made horticultural environments) could potentially be occupied by the target fruit fly species because of their polyphagous host range (De Meyer *et al.*, 2008).

2.3 Ecological niche modeling

The comparison of prediction performance were made among three methods, *i. e.*, BIOCLIM (Busby, 1991), DOMAIN (Carpenter *et al.*, 1993) and Maxent (Phillips *et al.*, 2006). The final prediction maps of three fruit fly species were produced based on species occurrence data using the superior one. In each method, some rules or mathematical algorithms were used to define the ecological niche of the species based on the distribution records of the fruit fly species in the multidimensional environmental space. Once the species niche is defined, its projection into the geographical space produces a predictive map.

BIOCLIM defines the ecological niche of a species as the bounding hyper-box that encloses all the records of the species in the climatic space (Busby, 1991). Thus, it creates a rectilinear ‘envelope’ in the environmental space, defined by the most extreme (minimum and maximum) records of the species on each environmental variable. To reduce the sensitivity of model predictions to outliers, the species records are sorted along each variable and only the records that lie within a certain percentile range of the data are used for model construction.

DOMAIN uses a point-to-point similarity metric (based on the Gower distance) to assign a value of habitat suitability to each potential site based on its proximity in the environmental space to the closest (most similar) occurrence location (Carpenter *et al.*,

1993). A threshold value of suitability can then be selected to determine the boundaries of the ecological niche. In contrast to all previous methods, environmental envelopes defined by DOMAIN are not necessarily continuous in the environmental space. BIOCLIM and DOMAIN were launched on DIVA-GIS Version 5.4.0.1 (Hijmans *et al.*, 2001), which was downloaded from <http://www.diva-gis.org/>. We followed recent research papers to tune the parameters and thresholds of the two models, and used the True-False option, to place pixels into one of two categories.

In this study, for BIOCLIM, two percentile ranges were applied: 90% (disregarding 5% of the values on each side), 95% (disregarding 2.5% of the values on each side); for DOMAIN, two thresholds were chosen: 0.95 and 0.97 (Hernandez *et al.*, 2006; Beaumont *et al.*, 2007; Tsoar *et al.*, 2007).

Maxent is a machine learning method, which estimates the distribution of a species by finding the probability distribution of maximum entropy (*i.e.* that is closest to uniform) subject to constraints representing our incomplete information about the distribution. The constraints are the expected value of each environmental variable should match its average over sampling locations derived from environmental layers. The model evaluates the suitability of each grid cell as a function of environmental variables at that cell. The suitability values provided by Maxent range from 0 (unsuitable habitat) to 100 (optimal habitat). Maxent version 3.2.19 downloaded from <http://www.cs.princeton.edu/~schapire/maxent/> was used to do the predictions. We followed recently published best practice approaches to tune its parameters (Phillips *et al.*, 2006; Phillips and Dudik, 2008). The default settings were used for most Maxent parameters (Auto Features, Random Test percentage; 0, Regularization multiplier; 1, Convergence threshold; 0.00001, Logistic output format), and with a few exceptions (maximum iterations; 1 000). In addition, Maxent has a built-in jackknife option through which the importance of individual environmental data layers can be estimated (Fig.5).

2.4 Accuracy assessment and performance comparison methods

Maxent predictions (Maximum training sensitivity plus specificity logistic threshold) were compared with two thresholds of predictions of BIOCLIM and DOMAIN respectively in this study. A non-parametric test (Mann-Whitney *U*) was used to determine whether one prediction is significantly better than another in the same sample sets. Three methods were used to assess the prediction accuracy of models: First, the area under the curve (AUC) in receiver operating characteristic analyses (ROC) (Elith, 2002; Phillips *et al.*, 2006); Second, the Cohen's kappa values from an error matrix (Cohen, 1960); Third, the True

Skill Statistic (TSS) (Allouche *et al.*, 2006). All layer processing was done on ArcGIS 9.2 Desktop (ESRI Inc., 2006).

Receiver operating characteristic (ROC) plot was selected as the first method to evaluate the accuracy of generated distributions. A curve that maximizes sensitivity for low values of the false-positive fraction is considered a good prediction and is quantified by calculating the AUC. The AUC can be used as a measure of overall performance of the model and has values usually ranging from 0.5 (random) to 1.0 (perfect discrimination), if the values below this range indicating a prediction that is worse than random (Engler *et al.*, 2004). The main advantage of ROC analysis is the area under the ROC curve (AUC) providing a single measure of model performance, independent of any particular choice of threshold. While generally used when data are available, ROC plots can also be created with presence and background absence data. In this study, Hawth's Analysis Tools for ESRI's ArcGIS (v3.x) software (Beyer, 2007) was implemented to randomly generate 100 background observations in global; and we put them into the ROC curve procedure in place of absence points to accompany the presence observations (Hernandez *et al.*, 2006). The AUC values are calculated in SPSS 16.0 (SPSS Inc., 2007).

Cohen's kappa is the most popular measure for the accuracy of presence-absence predictions (Loiselle *et al.*, 2003; Petit *et al.*, 2003; Rouget *et al.*, 2004; Segurado and Araújo, 2004). The kappa statistic ranges from -1 to +1, where +1 indicates perfect agreement and values of zero or less indicate a performance no better than random. Both commission and omission errors are accounted for in one parameter.

Although both ROC/AUC and Cohen's kappa are widely used in prediction accuracy assessment, several studies have criticized their associated shortcomings (Allouche *et al.*, 2006; Lobo *et al.*, 2008). So true skill statistic (TSS) was adopted as the alternative measure of prediction accuracy assessment which is not affected by prevalence and the size of the validation set, and two methods of equal performance have equal TSS scores (Allouche *et al.*, 2006). Like kappa, TSS takes into account both omission and commission errors, and success as a result of random guessing, and ranges from -1 to +1, where +1 indicates perfect agreement and values of zero or less indicate a performance no better than random.

2.5 Jackknife test

A jackknife test was then performed for each species to assess the relative contribution of the individual environmental variables in the prediction procedure. The test was performed as follows: (i) the full complement of 19 environmental variables was used

to perform the prediction as a baseline for comparisons; (ii) each environmental variable was eliminated systematically to provide the 19 possible combinations of 18 environmental variables each, and the 19 possible combinations were used to analyze the importance of each missing variable; (iii) similarly, but in the opposite way, each environmental variable was included systematically in single-variable analysis (Peterson and Cohoon, 1999; Levine *et al.*, 2007).

Although jackknife tests can be performed by hand, we used Maxent to conduct them because the whole procedure was automatically executed, and the results were intuitive and easy to interpret (Phillips, 2006; Phillips *et al.*, 2006).

2. 6 Application procedures for predictive performance comparison and potential distribution prediction

The whole approach comprised of six main steps (Fig. 1): (1) Dividing the occurrence samples of *D. bivittatus* into two parts stochastically: 80% and 20% parts by a SAS program (SAS Institute Inc., 2003); (2) Using the 80% of *D. bivittatus* occurrence points to create the prediction maps by BIOCLIM, DOMAIN and Maxent; (3) Testing the prediction accuracy of three models by ROC/AUC, Cohen's kappa and TSS; (4) Repeating Step 1 to Step 3 ten times; (5) Finding out the superior model; (6) Using the all occurrence samples of each fruit fly species to predict their habitat areas by the superior model. Since our goal is to predict global potential habitat areas, we test model performance only for *D. bivittatus* (80% of the records), which has the most occurrence records (De Meyer *et al.*, 2008). In total, each partition for *D. bivittatus* held 276 training points and 69 test points. Ten random partitions were made to assess the average behavior of models.

3 RESULTS

3.1 Performance comparison

Fig. 2 indicates that Maxent has the best performance (with the highest AUC, Kappa and TSS); DOMAIN has the median performance; and BIOCLIM with both thresholds performed comparatively poorly. DOMAIN with 0.97 threshold has the closest AUC value with Maxent, and the difference is not significant (two-tailed Mann-Whitney *U* test; P of AUC = 0.65); DOMAIN with 0.95 threshold has the closest Kappa and TSS value with Maxent, and the differences are both Extremely significant (two-tailed Mann-Whitney *U* test; P of Kappa = 0.000, P of TSS = 0.000). Based on the performance comparisons mentioned above, we used the predictions of Maxent to analyze the potential distributions of the three species of fruit flies, which present each pixel's probability in the form of continuous value (corresponding logistic value).

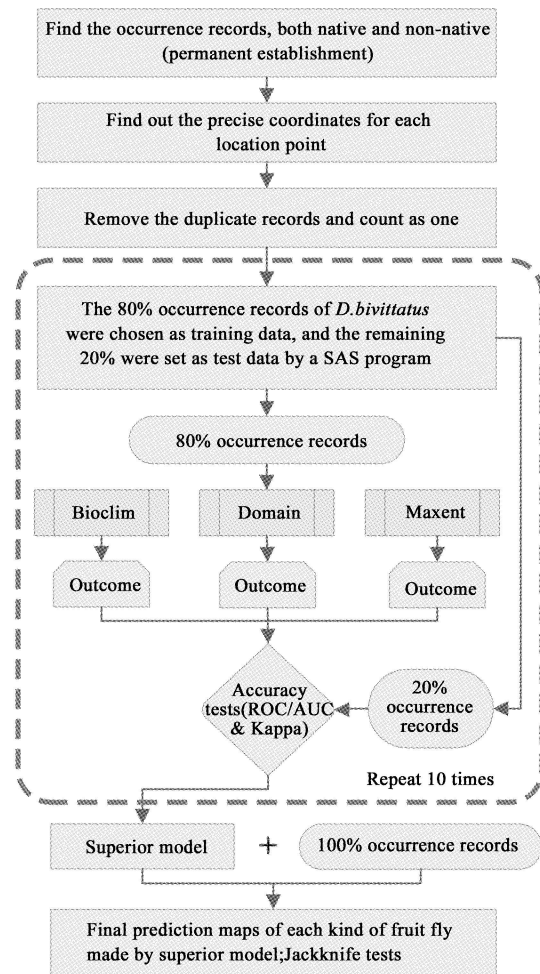


Fig. 1 Procedures for the predictive performance comparison among BIOCLIM, DOMAIN and Maxent

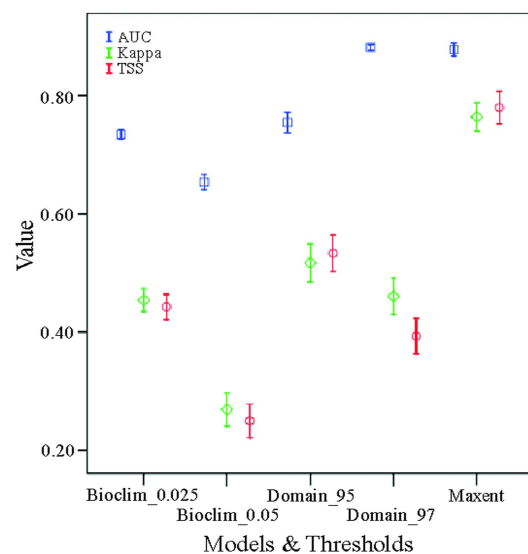


Fig. 2 AUC, Kappa and TSS value comparisons among BIOCLIM, DOMAIN and Maxent in different thresholds

The error bars represent 95% confidence interval for mean. On X axis, the figures beside 'BIOCLIM' and 'DOMAIN' are their percentile ranges, the threshold of Maxent is the maximum training sensitivity plus specificity logistic threshold.

3.2 Prediction of potential distribution areas

Fig. 3 shows the potential distribution areas of three fruit fly species in global area made by Maxent. Although the difference of global area distribution coverage of three fruit fly species is not significant, *D. ciliatus* has the comparatively widest distribution areas among three fruit fly species, the latter is *D. bivittatus*; and *D. vertebratus* has the smallest distribution areas. For *D. ciliatus*, the long island of USA, wide areas of Mexico, Bahamas, Cuba, Central America, South America, Africa, Saudi Arabia, Yemen, Oman, south Iran and Sri Lanka, some areas of Europe (costal areas of Portugal, Spain, Italy and Greece), Yemen, Jordan, Oman, Pakistan, India, Southeast Asia and Australia (coastal areas) are all its

potentially suitable habitat areas (Fig. 3: B). In addition, it is the only species among three species that has the potential distribution in coastal areas around the Mediterranean Sea and Southern Europe. For *D. bivittatus*, the long island of USA, Mexico, Bahamas, Cuba, Central America, middle of South America, Africa, Yemen, Sri Lanka, some areas of Pakistan, India, Southeast Asia and Australia (coastal areas) are its suitable habitat areas (Fig. 3: A). For *D. vertebratus*, the long island of USA, wide areas of Mexico, Bahamas, Cuba, Central America, South America, sub-Saharan Africa, Yemen, Sri Lanka, some areas of Pakistan, India, Southeast Asia and Australia (coastal areas) are its suitable habitat areas (Fig. 3: C).

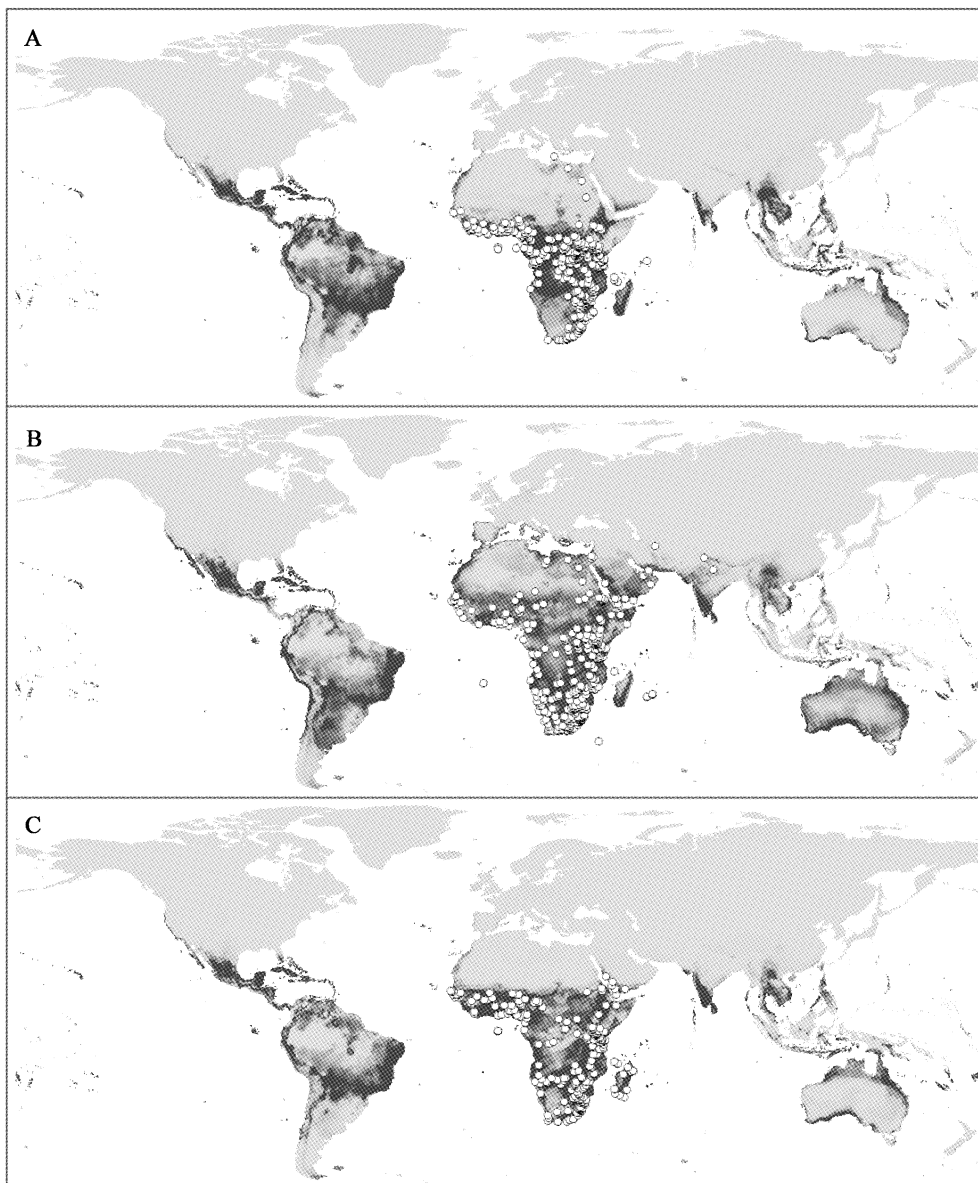


Fig. 3 Global potentially suitable areas of three fruit fly species predicted by Maxent and presented each pixel's probability in the form of continuous value (Corresponding logistic value)

A; *D. bivittatus*; B; *D. ciliatus*; C; *D. vertebratus*. The darker color of a pixel indicates the higher confidence in fruit fly presence. White dots on maps represent the locations and distributions of the three fruit fly species.

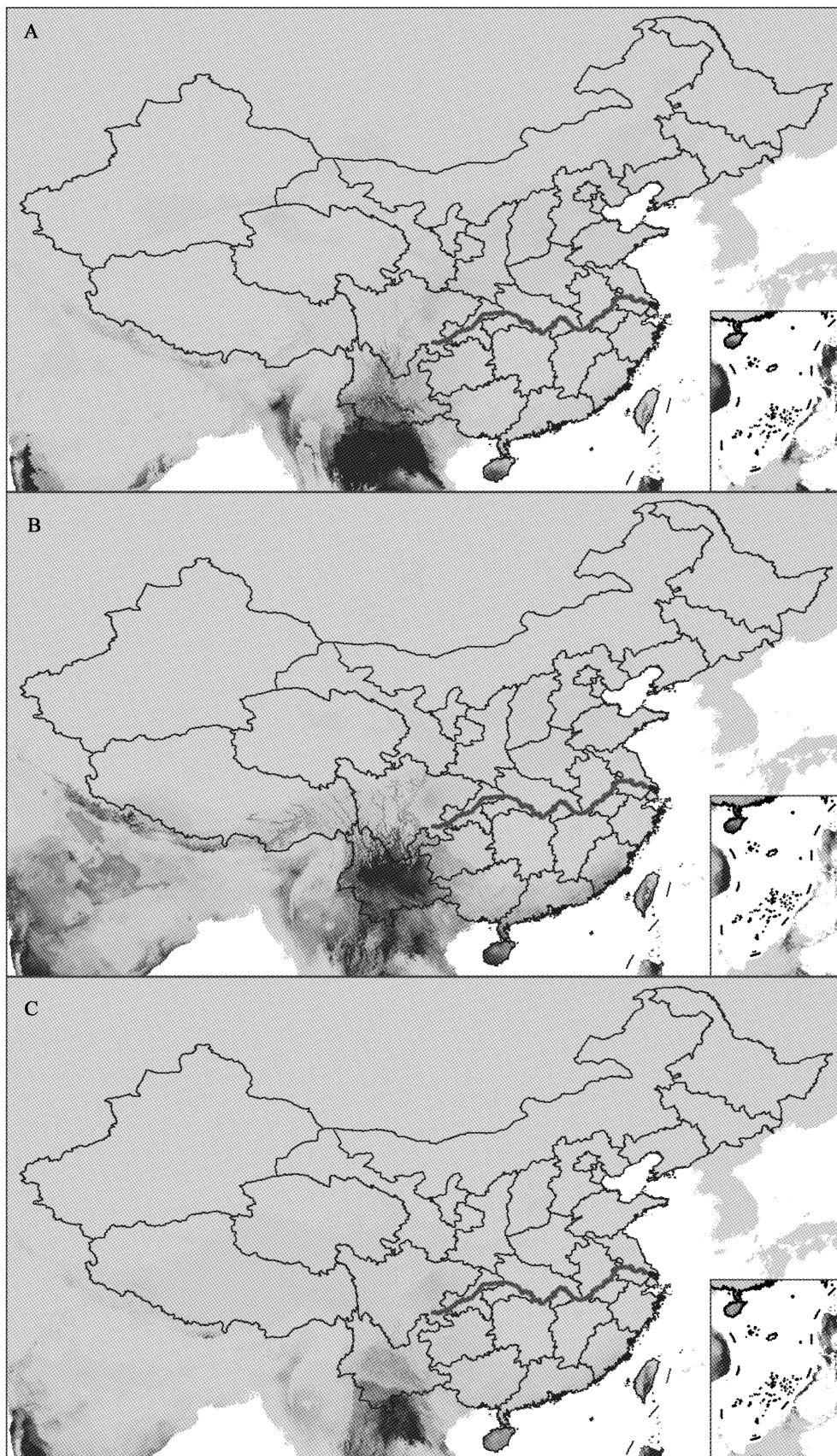


Fig. 4 Potential distributions of three fruit fly species in China predicted by Maxent and presented each pixel's probability in the form of continuous value (Corresponding logistic value)

A; *D. bivittatus*; B; *D. ciliatus*; C; *D. vertebratus*. The darker color of a pixel indicates the higher confidence in fruit fly presence, and bold lines on maps represent the Yangtze River.

Fig. 4 and Fig. 5 reveal that large areas of Yunnan and Hainan provinces of China are very habitable for all three fruit fly species, southern part of Guangdong and Taiwan are their habitat areas too. It is

indicated that *D. ciliatus* has the widest potential distribution areas in China, including southern part of Sichuan, Guizhou and Tibet. The coastal areas of southern China are its suitable areas.

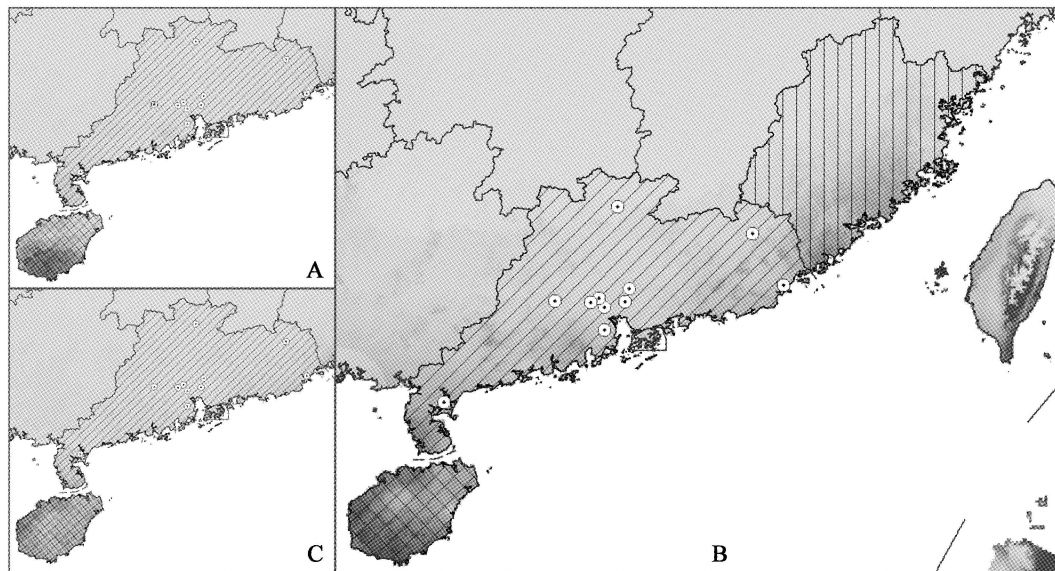


Fig. 5 Potential habitat areas of three fruit fly species in Guangdong predicted by Maxent and presented each pixel's probability in the form of continuous value (Corresponding logistic value)

A: *D. bivittatus*; B: *D. ciliatus*; C: *D. vertebratus*. The darker color of a pixel indicated the higher confidence in fruit fly presence. The areas covered by 10% simple hatch with 45° angle is Guangdong province, covered by 10% simple hatch with 90° angle is Fujian province, and covered by 10% crosshatch is Hainan province. White dots on the map are important cities of Guangdong where entry-exit ports are located (AQSIQ, 2005).

3.3 Jackknife analysis

Fig. 6 indicates that results of jackknife tests of three fruit fly species are similar with each others on the whole, and five variables have notable influences on distribution prediction. These are Isothermality, Mean Temperature of Coldest Quarter, Min Temperature of Coldest Month, Temperature Annual Range, and Temperature Seasonality. Meanwhile, five variables show less contribution to the prediction since they achieve few gains. These are Mean Diurnal Range, Precipitation Seasonality, Precipitation of Coldest Quarter, Precipitation of Driest Month, and

Precipitation of Driest Quarter.

For *D. bivittatus*, *D. ciliatus*, and *D. vertebratus* in that order, the environmental variables with highest gain when used in isolation are Temperature Seasonality, Isothermality and Isothermality, respectively (the longest blue bars in Fig. 6: A, B, C).

In addition, the light blue bars in Fig. 6 reveal that for all the three kinds of fruit fly, no variable has more useful information that is not already contained in the others, because omitting each one in turn didn't decrease the training gain much.

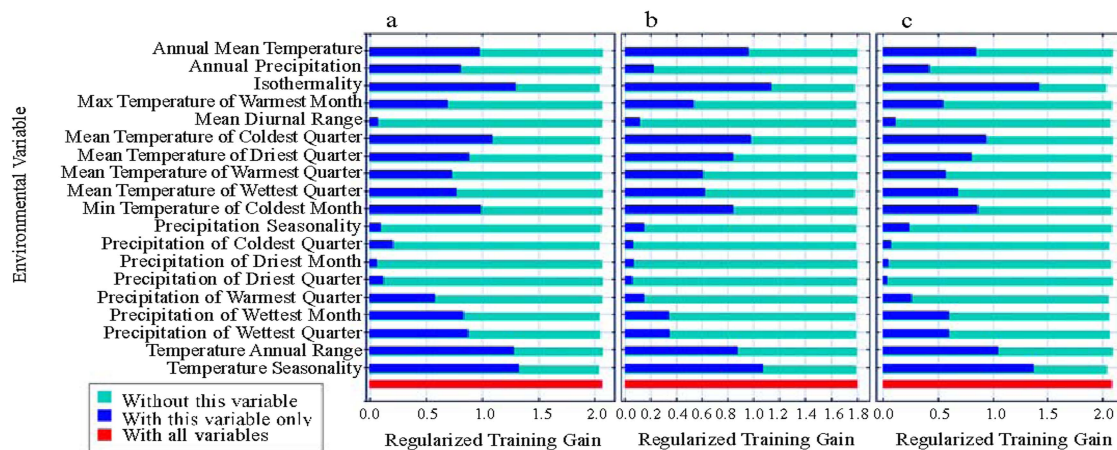


Fig. 6 Jackknife tests for environmental variable significance performed by Maxent
A: *D. bivittatus*; B: *D. ciliatus*; C: *D. vertebratus*.

4 DISCUSSION

Performance comparison of ecological niche models in this study indicates that Maxent has the best performance among the three models (Fig. 2); and in former similar studies, Maxent turned out to be the best when compared with other methods (ENFA, GARP, Maxent and PCA), which makes it a very promising method (Segurado and Araújo, 2004; Graham and Hijmans, 2006; Tsoar *et al.*, 2007). Besides, there are two additional advantages make it outstanding. First, Maxent can be easily interpreted by human experts, a property of great practical importance. Its construction idea is to estimate the target distribution by finding the distribution of maximum entropy (*i. e.* that is closest to uniform) subject to the constraint that the expected value of each feature under this estimated distribution matches its empirical average. This turns out to be equivalent, under convex duality, to finding the maximum likelihood Gibbs distribution. For species distribution modeling, the occurrence localities of the species serve as the sample points, the geographical region of interest is the space on which this distribution is defined, and the features are the environmental variables (Phillips *et al.*, 2004). Second, it has a high success rates and statistical significance with sample sizes as low as five occurrence records (Pearson *et al.*, 2007).

The prediction accuracy of GARP and Maxent is greatly influenced by exactly which observations are included; the importance of each record depends on whether that locality represents a unique environment not represented by the other sample points (Pearson *et al.*, 2007). In this research, the distribution records of each fruit fly species covered all its native-range distribution countries and included the important non-native distribution areas, which represent a wide variety of environment conditions and make the final outcomes be of high accuracy.

The final prediction outcomes indicate that the risk of three fruit fly species permanently established in Guangdong exists but low (Figs. 4 and 5). *D. ciliatus* has the widest potential distribution areas in China, followed by *D. bivittatus* and *D. vertebratus*. Wide areas of Fujian are suitable for *D. ciliatus*. In addition, the neighbor countries of China (*e. g.*, Laos, Thailand and Vietnam), where are also suitable for the three fruit fly species, should not be ignored.

Based on jackknife analysis, temperature and its variation influence the distribution patterns of three fruit fly species significantly both in global and restricted regions (Fig. 6), and which is consistent with former research (Vayssieres *et al.*, 2008).

Temperature is a very important abiotic factor affecting immature development of *Dacus* spp. Knowledge of relationships between developmental time and temperature is essential for predicting population fluctuations in the field. According to Vayssieres (2008), there was a strong linear regression model between temperature and development rate of *D. ciliatus*. Temperature thresholds (T , °C) and thermal constants (K , degree days) of the different immature stages of *D. ciliatus* (egg, larva, pupa and total developmental stage in order) on pumpkin are: 10.56°C and 26 degree days, 10.23°C and 90 degree days, 11.20°C and 207 degree days, and 11.04°C and 323 degree days, respectively.

Since 1980, the African people have emerged into China (Guangdong mostly) at a high speed which keeps raising intensely per year. In 2005, the number of incoming travelers from Africa at Guangzhou is up to 31 766, and the average increasing rate per year is 37.9%. Nowadays, the population probability of African in Guangzhou and Shenzhen ranges from 1/50 to 1/25 (Li *et al.*, 2008; Pan *et al.*, 2008). It is very possible that *D. bivittatus*, *D. ciliatus* and *D. vertebratus* introduced into China with incoming travelers.

Many biological invasions result from human activity. They are mainly considered as one of the consequences of increasing international trade. Once an exotic fruit fly species is established, it cannot, in most cases, be eradicated. Attempts at management are often not particularly effective (Vayssieres *et al.*, 2008). Forecasting tephritid invasion is a major current issue for both tropical and temperate countries. In all cases, environmental factors, especially temperature, are likely to affect the invasion process. Also, these results are useful for improving laboratory-rearing methods and for building simulation models to predict fruit fly population dynamics.

The process of invasive species entering and establishing in a new region is complex, requiring about six barriers according to Richardson (2004): A, main geographical barriers; B, local environmental barriers; C, reproductive barriers; D, dispersal barriers; E, environmental barriers (disturbed habitats); F, environmental barriers (undisturbed habitats). In this study, potential habitat areas in China (barrier B) for the three species of fruit flies were estimated. We should also investigate potential competition of the invasive exotic species with indigenous species or with other introduced species, which is considered an important part of barriers C and D (Duyck *et al.*, 2004; De Meyer *et al.*, 2008). Today *Bactrocera dorsalis* (Hendel) and *Bactrocera cucurbitae* (Diptera: Tephritidae) are two most serious fruit fly species in China, and caused enormous

damage to fruit industry every year. They are widely distributed in southern provinces of China (Hou and Zhang, 2005; Zhou *et al.*, 2007). Further relevant studies should concentrated on the competition between invasive species and indigenous species based on concept of propagule pressure and Allee effects (Drake *et al.*, 2005; Lockwood *et al.*, 2005; Drake and Lodge, 2006).

ACKNOWLEDGMENTS This work was supported by National Basic Research Program of China (973 Program, 2002CB111405), the National Nature Science Foundation of China (30471162, 30671394), and the National Key Technologies Research and Development Program of China (2006BAD08A13).

References

- Allouche M, Tsoar A, Kadmon R, 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43: 1 223–1 232.
- AQSIQ, 2005. Plant inspection and quarantine. In: AQSIQ ed. Manual of Inspection and Quarantine. AQSIQ, Beijing. 80–81. [国家质量监督检验检疫总局, 2003. 植物检验检疫分册. 见: 国家质量监督检验检疫总局主编. 检验检疫工作手册. 北京: 国家质量监督检验检疫总局. 80–81]
- Beaumont LJ, Pitman AJ, Poulsen M, Hughes L, 2007. Where will species go? Incorporating new advances in climate modelling into projections of species distributions. *Global Change Biology*, 13: 1 368–1 385.
- Beyer H, 2007. Hawth's Analysis Tools for ESRI's ArcGIS, Beyer H [Online]. Available from: <http://www.spatial ecology.com/htools>.
- Busby JR, 1991. BIOCLIM: a bioclimate analysis and prediction system. In: Margules CR, Austin MP eds. Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne. 64–68.
- CABI, 2005. Crop Protection Compendium. CAB International, Wallingford.
- Carpenter G, Gillison AN, Winter J, 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, 2(6): 667–680.
- Cohen J, 1960. A coefficient of agreement of nominal scales. *Educational and Psychological Measurement*, 20: 37–46.
- De Meyer M, Robertson MP, Peterson AT, Mansell MW, 2008. Ecological niches and potential geographical distributions of Mediterranean fruit fly (*Ceratitis capitata*) and Natal fruit fly (*Ceratitis rosa*). *Journal of Biogeography*, 35: 270–281.
- Drake JM, Baggenstosp, Lodge DM, 2005. Propagule pressure and persistence in experimental populations. *Biology Letters*, 1(4): 480–483.
- Drake JM, Lodge DM, 2006. Allee effects, propagule pressure and the probability of establishment: Risk analysis for biological invasions. *Biological Invasions*, 8: 365–375.
- Duyck PF, David P, Quilici S, 2004. A review of relationships between interspecific competition and invasions in fruit flies (Diptera: Tephritidae). *Ecological Entomology*, 29: 511–520.
- Elith J, 2002. Quantitative methods for modeling species habitat: Comparative performance and an application to Australian plants. In: Ferson S, Burgman M eds. Quantitative Methods for Conservation Biology. Springer-Verlag, New York.
- Engler R, Guisan A, Rechsteiner L, 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, 41: 263–274.
- EPPO, 2007. Plant Quarantine Data Retrieval System. EPPO, France, Paris.
- ESRI Inc., 2006. ArcGIS 9.2 Desktop. Environmental Systems Research Institute, Inc., Redlands, CA.
- Graham CH, Hijmans RJ, 2006. A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, 15: 578–587.
- Hernandez PA, Graham CH, Master LL, Albert DL, 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29: 773–785.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25: 1 965–1 978.
- Hijmans RJ, Guarino L, Cruz M, Rojas E, 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter*, 127: 15–19.
- Hou BH, Zhang RJ, 2005. Potential distributions of the fruit fly *Bactrocera dorsalis* (Diptera: Tephritidae) in China as predicted by CLIMEX. *Acta Ecologica Sinica*, 25(7): 1 569–1 574. [侯柏华, 张润杰, 2005. 基于 CLIMEX 的桔小实蝇在中国适生区的预测. 生态学报, 25(7): 1 569–1 574]
- Levine RS, Peterson AT, Yoritani KL, Carroll D, Reynolds MG, 2007. Ecological niche and geographic distribution of human Monkeypox in Africa. *PLoS One*, 2(1): 1–7.
- Li ZG, Xue DS, Lyons M, Brown A, 2008. The African enclave of Guangzhou: a case study of Xiaobeilu. *Acta Geographica Sinica*, 63(2): 207–218. [李志刚, 薛德升, Michael Lyons, Alison Brown, 2008. 广州小北路黑人聚居区社会空间分析. 地理学报, 63(2): 207–218]
- Liang F, Zhao JP, Hu XN, Liang GQ, Wu JJ, 2007. Aquarantime pest: *Dacus (Dacus) bivittatus* was intercepted and captured firstly in Guangzhou airport. *Chinese Bulletin of Entomology*, 44(6): 609–809. [梁帆, 赵菊鹏, 胡学难, 梁广勤, 吴佳教, 2007. 我国首次截获检疫性害虫——葫芦蓟实蝇. 昆虫知识, 44(6): 609–809]
- Liang F, Zhao JP, Liang GQ, Gu JL, Hu XN, Wu JJ, 2006. *Dacus ciliatus* Low was intercepted and captured firstly in Guangzhou airport. *Plant Quarantine*, 120(14): 261–262. [梁帆, 赵菊鹏, 梁广勤, 古菊兰, 胡学难, 吴佳教, 2006. 我国截获小南瓜实蝇. 植物检疫, 120(14): 261–262]
- Liang GQ, Liang GZ, Lin M, Liang F, 1993. Fruitfly species and their management and control. Guangdong Science & Technology Press, Guangzhou. 74–81. [梁广勤, 梁国真, 林明, 梁帆, 1993. 实蝇及其防除. 广州: 广东科技出版社. 74–81]
- Lobo JM, Jiménez-Valverde A, Real R, 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17: 145–151.
- Lockwood JL, Casseyp, Blackburn T, 2005. The role of propagule pressure in explaining species invasions. *Trends in Ecology and Evolution*, 20(5): 223–228.
- Loiselle BA, Howell CA, Graham CH, Goerck JM, Brooks T, Smith KG, Williams PH, 2003. Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, 17: 1 591–1 600.
- Pan XL, Chen C, Yang DZ, 2008. The chocolate city: Africans chasing dreams in Guangzhou. *Caiju Jingwei*, (7): 102–103. [潘晓凌, 陈晨, 杨大正, 2008. 巧克力之城——非洲人寻梦广州. 财富经纬, (7): 102–103]
- Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT, 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34: 102–117.
- Peterson AT, Cohoon KP, 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling*, 117: 159–164.
- Petit S, Chamberlain D, Haysom K, Pywell R, Vickery J, Warman L, and DA, Firbank L, 2003. Knowledge-based models for predicting species occurrence in arable conditions. *Ecography*, 26: 626–640.
- Phillips SJ, 2006. A brief tutorial on Maxent. [Online]. Available from: <http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc>.
- Phillips SJ, Anderson RP, Schapire RE, 2006. Maximum entropy

- modeling of species geographic distributions. *Ecological Modelling*, 190: 231–259.
- Phillips SJ, Dudik M, 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31: 161–175.
- Phillips SJ, Dudik M, Schapire RE, 2004. A maximum entropy approach to species distribution modeling. In: Proceedings of the 21st International Conference on Machine Learning, Banff, Canada.
- Richardson DM, Van Wilgen BW, 2004. Invasive alien plants in South Africa: How well do we understand the ecological impacts? *South African Journal of Science*, 100: 45–52.
- Rouget M, Richardson DM, Nel JL, Maitre DCL, Egoh B, Mgid T, 2004. Mapping the potential ranges of major plant invaders in South Africa, Lesotho and Swaziland using climatic suitability. *Diversity and Distributions*, 10: 475–484.
- SAS Institute Inc., 2003. SAS for Windows 9.1.3 Service Pack 2. SAS Institute Inc., Cary, NC, USA.
- Segurado P, Araújo MB, 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography*, 31: 1 555–1 568.
- SPSS Inc., 2007. SPSS 16.0 for Windows. SPSS Inc., Chicago, IL.
- Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R, 2007. A comparative evaluation of presence only methods for modelling species distribution. *Diversity and Distributions*, 13: 397–405.
- Vayssières JF, Carel Y, Coubes M, Duyck PF, 2008. Development of immature stages and comparative demography of two cucurbit-attacking fruit flies in Reunion Island: *Bactrocera cucurbitae* and *Dacus ciliatus* (Diptera: Tephritidae). *Environmental Entomology*, 37(2): 307–314.
- Zhou GL, Chen C, Ye J, Hu BS, Liu FQ, 2007. Predicting potential ecological distribution of *Bactrocera dorsalis* in China using GARP ecological niche modeling. *Acta Ecologica Sinica*, 27(8): 3 362–3 369. [周国梁, 陈晨, 叶军, 胡白石, 刘凤权, 2007. 利用 GARP 生态位模型预测桔小实蝇 (*Bactrocera dorsalis*) 在中国的适生区域. 生态学报, 27(8): 3 362–3 369]

基于最大熵值法生态位模型 (Maxent) 的 三种实蝇潜在适生性分布预测

李白尼¹, 魏 武¹, 马 骏², 张润杰^{1,*}

(1. 中山大学有害生物控制与资源利用国家重点实验室/昆虫学研究所, 广州 510275;

2. 广东出入境检验检疫局检验检疫技术中心, 广州 510623)

摘要: 本研究首先对 3 种重要生态位模型 BIOCLIM, DOMAIN 和 Maxent(基于最大熵值原理模型)的分布预测精确度进行了分析和比较,再结合分布点记录以及一系列环境数据图层对 3 种重要外来入侵性检疫害虫(葫芦寡鬃实蝇 *Dacus bivittatus*、埃塞俄比亚寡鬃实蝇 *D. ciliatus* 和西瓜寡鬃实蝇 *D. vertebratus*)的潜在适生性分布区域进行了预测和分析。在模型预测精确度的比较过程中,3 种评估指标(ROC/AUC, Kappa, TSS)均显示 Maxent 拥有最好的预测结果和最好的运行性能。由 Maxent 对葫芦寡鬃实蝇、埃塞俄比亚寡鬃实蝇和西瓜寡鬃实蝇的预测结果显示,这 3 种实蝇在中美洲、南美洲、东南亚和澳大利亚沿岸的广大地区在总体上具有相似的分布区域。相对而言,埃塞俄比亚寡鬃实蝇在全球范围具有最为广泛的分布区域,除前述地区外,其潜在适生区还包括地中海沿岸、沙特阿拉伯、也门、安曼和伊朗南部的大片地区,这也意味着在 3 种寡鬃实蝇中,它能忍受变化幅度最广的生态、环境条件。在中国,云南和海南都极适宜于 3 种实蝇的生存,同时广东南部及台湾的部分地区也是它们的潜在适生区。基于 Maxent 的预测结果显示,相对而言,埃塞俄比亚寡鬃实蝇在中国范围也具有最为广泛的分布区域,除前述省份和地区外,四川、贵州和西藏的南部部分地区以及中国南部的部分沿海地区,也都是它的潜在适生区。综合所得出的预测结果,3 种寡鬃实蝇从境外传入广东并在此定殖的风险可能性是实际存在的。Jackknife 分析显示,温度以及与此有关的环境因子对于 3 种实蝇在全球和局部地区的分布模式和分布情况都有极大的影响,并需要进一步的研究。

关键词: 葫芦寡鬃实蝇; 埃塞俄比亚寡鬃实蝇; 西瓜寡鬃实蝇; 生态位模型; 最大熵值; Maxent; BIOCLIM; DOMAIN
中图分类号: Q968.1 **文献标识码:** A **文章编号:** 0454-6296(2009)10-1122-10

(责任编辑: 袁德成)